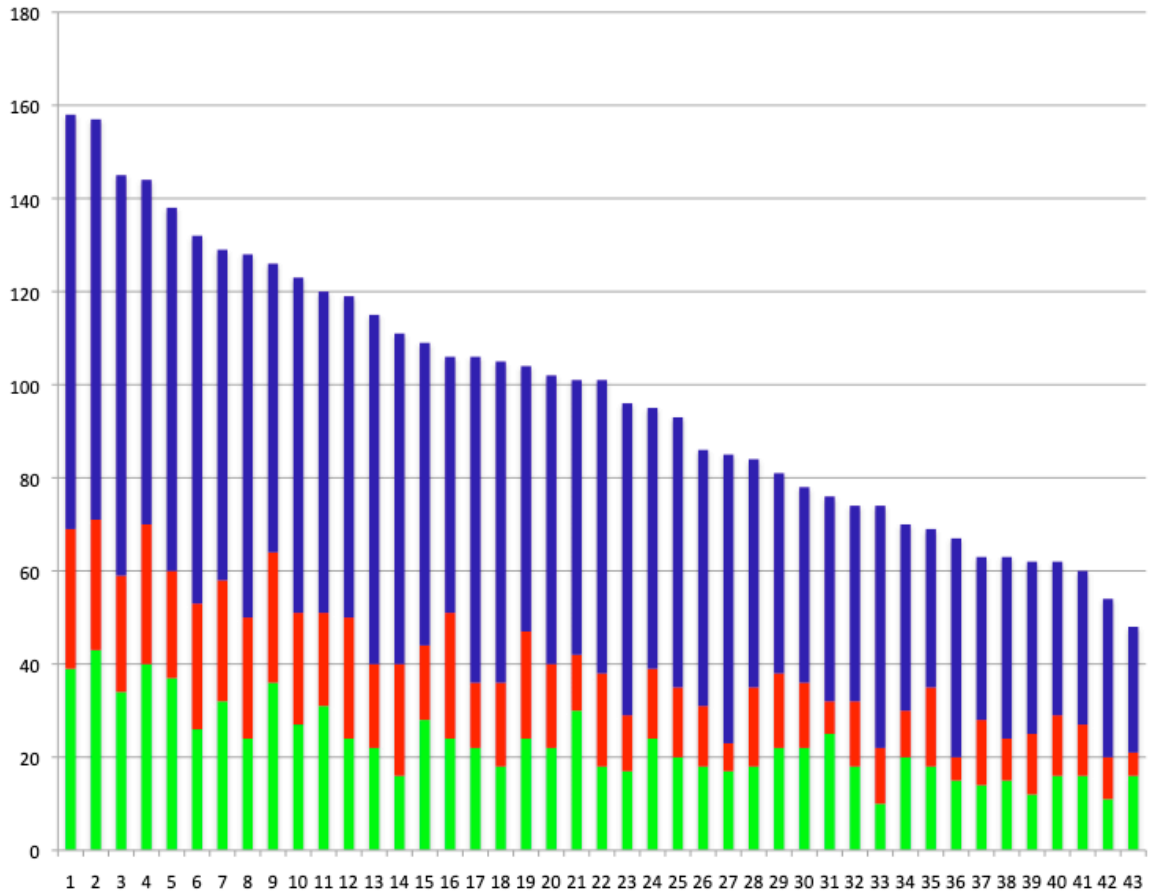
COS 109 Final Exam, Fall 2019

I graded this myself. The median grade was 101, with quartiles at 120 and 74. The median from the 2018 exam was 97, with quartiles at 113 and 76. The colors in the graphs below are for parts 1, 2 and 3, reading up from the bottom.



1. **(50 points, 2 each) Short Answers**. Circle the right answer or write it in the space provided.

(a) The I Ching (易经) is an ancient Chinese scheme for divination. It consists of a set of hexagrams like this ䷀, composed of six horizontal lines, each of which is either solid or broken. How many hexagrams are there in the I Ching?

64. 6 rows, each of which can be either solid or broken; hence 2^6.

(b) If **m** and **n** are positive integers, how many 1-bits (that is, bits whose value is 1) are there in the binary representation of $(2^m \times 2^n) - 1$?

m + n. This expression is 1 less than a power of two, and as discussed multiple times, that means that it's all 1s.

(c) For many web sites at Princeton, one logs in through a common OIT login page; after you present your netid and password successfully, you arrive, fully validated, at the application on the other side. In the usual *dramatis personae* of cryptography (Alice, Bob, et al), what role or character is OIT playing?

Trent, the trusted third party.

(d) The letters of T. S. Eliot to Emily Hale, written from 1930 to 1957, were donated to Princeton by Ms Hale with the stipulation that they remain sealed until 50 years after both parties had died. Eliot died in 1965, Hale died in 1969, and the letters were unsealed for scholarly research on January 2, 2020. Under current American law, in what year did or will Eliot's copyright on the letters expire?

2035, which is 70 years after Eliot's death.

(e) If Alice and Bob are discussing the significance of Etaoin Shrdlu, which of these is the most likely topic of their conversation?

compression **error detection** **machine translation** **passwords** **programming languages**

compression. These are the most common letters in normal English and thus would be represented by fewer bits in variable length encodings or compression techniques.

(f) In December 2019, researchers at McMaster University constructed the world's smallest gingerbread house, out of silicon, using an electron microscope. Length of the house: 10 micrometers. Suppose that the house was scaled down uniformly in all dimensions from an original version whose length was 10 centimeters. By what factor is the *volume* of the original house greater than the volume of the little house?

10^12. Each dimension is larger by 10^4 and there are 3 dimensions. Missed by many. Think volume!

(g) Famous Stanford computer scientist Don Knuth recently said, "I have just celebrated my 10000th birthday (in base 3)." How old is Don (in decimal)?

81. 10000 in base 3 is 3^4. Missed by quite a few.

(h) "The machine has a finite number of configurations. It has a tape divided into squares each capable of bearing a symbol. At any moment just one symbol is being scanned. The behavior of the machine is determined by the configuration and the scanned symbol. In some configurations, the machine writes a new symbol on the scanned square; in other configurations it erases the scanned symbol. The machine may also change the square which is being scanned, but only by shifting it one place to the right or left." Who designed the computer described in this paraphrase of the original paper?

Charles Babbage **John Hennessy** **Grace Hopper** **Ada Lovelace**
Gordon Moore **Alan Turing** **John von Neumann**

Turing. Recall the video of a modern incarnation of a Turing machine?

(i) "New Jersey's largest hospital system paid hackers a ransom after a ransomware attack disrupted its services in December 2019." Ransomware encrypts files on a victim's computer, and the victim has to pay the bad guys for a password that will decrypt them. If you were (just hypothetically) a bad guy, which of these mechanisms would be most suitable for implementing your ransomware?

AES **DES** **MD5** **ROT13** **RSA** **SHA-256** **TOR**

AES. DES is too easy to crack; RSA is public-key, operationally too complicated, and likely too slow.

ROT13 wouldn't be much of a threat.

(j) Suppose that I have an old program, written in a long-obsolete assembly language, that originally ran on a computer that no longer exists anywhere. Which of these software components would I need so I could run the old program on a current computer?

assembler **compiler** **simulator** **assembler + compiler**

assembler + simulator **compiler + simulator** **all three**

assembler+simulator. This is an assembly language program, so no compilers are involved.

(k) Suppose an old hard disk has a diameter of 3 inches and a uniform density of magnetic spots over its whole surface. If there are 100 million bits in a track one inch from the center, how many bits would there be in a track at the outer edge? (Tracks are all the same width.)

150 million. The outer edge is 1.5 inches from the center: the diameter is 3 inches.

(l) The Great Cannon is a Chinese cyber-warfare tool that "injects malicious Javascript into pages served from behind the Great Firewall. These scripts, potentially served to millions of users across the Internet, hijack the users' connections to make multiple requests against the targeted site." What kind of attack is this?

DDoS **DoS** **IoT** **MITM** **MS-DOS** **Trojan horse**

DDoS. Distributed denial of service; "distributed" because the attacks are coming from millions of sources all at once. Trojan horse was a popular wrong answer.

(m) A *NY Times* story (12/16/19) describes how companies as diverse as Netflix, Apple, Airbnb and General Electric "rent computing from _____ -- otherwise known as using the _____ -- instead of running their own systems." What is the company name and the kind of service.

Amazon, cloud. I accepted Microsoft and Google as alternate cloud providers, though Amazon does serve these specific companies at the moment.

(n) A *NY Times* article on cellphone tracking (12/19/19) says "Location data is also collected and shared alongside a mobile advertising ID, a supposedly anonymous identifier about 30 digits long that allows advertisers and other businesses to tie activity together across apps." Approximately how many bits long is that ID?

100. 10^30 is closest to 2^100.

(o) While playing with **`traceroute`**, I visited amazon.com, amazon.ca, amazon.de, amazon.fr and amazon.tv. What is the minimum number of different countries where these servers might be located?

1 **2** **3** **4** **5** **no way to tell**

1. As seen in experiments in class, servers can be anywhere, independent of country code.

(p) Harvard's CS50 course (loosely like COS 126 here) uses "a pairwise comparison algorithm" to examine problem set submissions to see if any pairs are suspiciously similar. How does the running time of this algorithm grow in proportion to N, the number of students in the class?

N^2. "Pairwise" should suggest comparing each submission to each other.

(q) Put these names into chronological order of when they made the contribution(s) that caused them to be mentioned in COS 109, by writing the numbers 1 through 5 on them.

Anthony Babington **Tim Berners-Lee** **Jeff Bezos** **Ada Lovelace** **Dennis Ritchie**

Babington, Lovelace, Ritchie, Berners-Lee, Bezos. Many people thought Ada came first, but Babington's broken crypto with Mary Queen of Scots was in 1586 or so; Ada is 1830-ish.

(r) Suppose we start a Towers of Hanoi game with **n** disks at the beginning of class. If it takes 1 minute to move one disk from one pin to another (they're very heavy, being made of solid gold), what's the largest value of **n** for which we could finish a game in an 80-minute class period?

6. 2^6 is 64, 2^7 is well beyond 80 minutes.

(s) What Princeton PhD's portrait hangs in the atrium of Lewis Library?

Turing. It took quite a while to arrive, but I kept mentioning it in class, including after it did show up.

(t) The page of RGB colors on the COS 109 web site includes 101 shades of gray called gray00 through gray100. What is the hex value of gray50 likely to be in the standard RGB representation?

7F7F7F. I also accepted 808080.

(u) Suppose that Quicksort encounters a really perverse data set: every single time the algorithm splits a group of items into two smaller groups, the split puts a single item in one group and all the other items in the other group. What is the likely running time of this variant as a function of **N**, the number of items to be sorted?

$\sqrt{N}$ **N** **N log N** **N^2** **2^N** **no way to tell**

N^2. At each step only element is moved, so the algorithm takes n + n-1 + n-2 + … steps.

(v) Suppose that one afternoon during an endless COS 109 lecture you use Safari to visit Amazon, Facebook and Google from your laptop, and Amazon sends you a cookie. Which of the following sites will receive that cookie later that day when you visit Amazon from your phone?

only Amazon **Amazon+Facebook** **Amazon+Google** **all three of them** **none of them**

None of them. Cookies are stored by your browser on your computer or phone, so the transactions are unrelated.

(w) According to *Advertising Age* (11/4/19), T-Mobile is suing Lemonade, an insurance company, over the latter's use of the color magenta, which T-Mobile claims it owns. What specific kind of intellectual property is at issue here?

Trademark. Recall our discussion of distinctive colors like Coke and fiberglass.

(x) In TV's long-running quiz show *Jeopardy*, in the category "Old Folks in Their Thirties," the clue was "Linus Torvalds is the father of this operating system used on cell phones & supercomputers." What's the answer?

Linux. Almost everyone got this gift.

(y) In the *Jeopardy* category "Let's Get I.T. On," the clue was "This programming language isn't a little worse than B minus; it's a 1980s improvement of a language called C." What's the answer?

C++. Almost as many. (Not C+; there is no such language.)

2. (30 points) Understanding Programs

(a) The following function is supposed to multiply two non-negative integers `m` and `n` by performing repeated addition. That is, `multiply(m,n)` should display the product `m * n`, so at the end `multiply(2,5)` should display `10`. Unfortunately, it doesn't work. Fix the errors. (This is a question about correct logic; don't worry about syntactic trivia, but make your corrected code clear.)

```
function multiply(m, n) {
    prod = 1;
    i = 1;
    while (i < n) {
        i = i - 1;
        prod = prod * m;
        alert("product is ", prod);
    }
}
```

```
prod = 0
i = 0
while (i < n) {
      prod = prod + m
      i = i + 1
}
alert("product is ", prod)
```

Not well done, I fear, with much confusion about how computation proceeds. "Repeated addition" should have suggested the need for an addition operator, and there have been quite a few examples of loops in problem sets and lectures.

(b) Once properly fixed, how does the running time of this algorithm depend on **m**?

logarithmic **linear** **m log m** **quadratic** **cubic** **exponential** **independent of m**

independent of m, as written here; the loop is controlled by n.

(c) How does the running time depend on **n**?

logarithmic **linear** **n log n** **quadratic** **cubic** **exponential** **independent of n**

linear in n. The loop is executed n times. These would be opposite if the loop were controlled by m.

(d) Here's a function in a Javascript-like language, with various parts identified by line numbers; the line numbers are not part of the function.

```
1:    function rangecheck(v, low, high)
2:    {
3:       if (v >= low and v <= high)
4:          return true
5:       else
6:          return false
7:    }
```

(i) Which part is the API for this function? Identify the line or lines by number.

Line 1

(ii) Which part is the implementation? Identify the line or lines by number.

Lines 3-6. The braces don't matter.

(iii) In a major court case, _____ and _____ will argue the copyrightability of APIs at _____ in March 2020. What belongs in the blanks?

Google, Oracle, Supreme Court.

(e) Suppose that some computational process produces an endless sequence of random integers between 1 and 100; any number is as likely as any other, so over a long enough period, any number will occur about as often as any other number. If the following fragment of Javascript-like code is used inside a loop to categorize 10,000 such numbers, about how many lines of each type of output would you expect to see?

```
if (num > 50)
   print "big"
else if (num > 30)
   print "middling"
else if (num > 5)
   print "small"
```

5000, 2000, 2500.

(f) Suppose that the Toy machine has a **MULT** instruction that multiplies the accumulator contents by some value and places the result back in the accumulator. What does this program print when given the sequence of numbers **1 –4 –2 5 –3 0** as its input?

```
BOT   GET           get a number from user, place it in accumulator
      IFZERO  TOP   if accumulator content is zero, go to location TOP
      STORE   MID   store accumulator content in location MID
      PRINT         print content of accumulator
      MULT    MID   multiply accumulator content by content of MID
      MULT    MID   multiply accumulator content by content of MID
      PRINT         print content of accumulator
      GOTO    BOT   go to location labeled BOT
TOP   STOP
MID   0             when execution begins, this location will contain 0
```

1 1 -4 -64 -2 -8 5 125 -3 -27. Prints the input numbers and their cubes.

3. (100 points, 5 each) Miscellaneous

(a) "eBay went from a system that allowed two billion auction items […] to one that can handle 18 billion billion." (*WSJ*, 5/6/09) The number of auction items is stored in an integer of some particular size.

(i) How many bytes did that integer probably have in the original system?

4. "2 billion" should suggest 2^31.

(ii) How many bytes did that integer probably have in the second system?

8. "18 billion billion" is about 2^4 * 10^9 * 10^9, or about 2^64

(iii) Which version used a sign bit in the representation of the integer value – the original, the second, neither, or both?

Original. If it were not using a sign bit, it would have overflowed at 4 billion, not 2. The second number uses all 64 bits (18 billion billion).

(b) *Base64 encoding* is a scheme that represents arbitrary binary data in a printable form. It's similar to hexadecimal, but uses 6-bit chunks instead of 4: each possible 6-bit pattern is encoded with a different letter or digit or other character.

(i) If an Ethernet packet is 1500 bytes long, how many characters would it take to write out the packet contents in Base64?

2000. Every 6 input bits becomes 8 output bits so the output expands by a third.

(ii) If instead of Base64, we use hexadecimal digits, how many characters would it take to write out the packet contents?

3000. Each byte takes two hex digits.

(c) *Steganography* is the art of concealing messages in apparently innocuous documents so that only the sender and the receiver are likely to know that there is a hidden message. For example, we might transmit a secret message by encoding its bits in the least significant bit of each byte of a standard RGB color image. That wouldn't change the visible colors by any perceptible amount but it would permit the transmission of hidden information.

(i) How many bytes would this technique hide in an 8 megapixel RGB image?

3 MB. One bit for every byte. 8 MP is 24 MB so 24 million bits or 3 MB.

(ii) Would this scheme work if the image were compressed with JPG? Yes or no.

No. JPG compression throws away information.

(d) Suppose that Annual Giving wants to store certain information about alumni donors in as few bits as possible. They plan to encode your graduation year (only two digits), your major (one of 30 departments), and your potential as a donor (in one of four categories). How many bits will this require per alum?

14. 7 bits for year (since there are 99 possibilities) + 5 for major (30) + 2 for potential (4). Not well handled, given how basic this is.

(e) Just before Christmas a few years ago, I got mail from someone in the university's PR department who wanted my help in sending out a tweet in binary, that is, with each ASCII character of the message spelled out as a sequence of 0's and 1's. (I am not making this up.) Suppose that the tweet is just the five characters "**`Xmas!`**". Write out the tweet in 8-bit binary.

0101 110 0110 1101 0110 001 0111 0011 0010 001. Another gift, mostly well done.

(f) Suppose we fill Lewis Library 138 from floor to ceiling with laptops. Ignoring desks, chairs, the irregular shape of the room and so on,

(i) Estimate how many laptops there would be. Be very clear about how you compute your estimate.

A million? Call the room 60 x 40 x 20 = 50,000 cubic feet. A laptop is 12 x 12 x 0.5 cubic inches, so there are about 20 per cubic foot. I took anything that was plausible. A fair number of people did not correctly convert cu ft to cu in: dividing by 12 is WRONG. Points off for excessive precision; given that you have no precise measurements for the room or laptop specifications, there can't be much more than one digit of precision in the results.

(ii) Estimate how many petabytes of memory there would be in total.

10 PB of RAM or 500 PB of disk? "Memory" is ambiguous, so I accepted either. Any reasonable sizes for either.

(iii) Estimate how many teraflops of processing power there would be.

1000 Tflops? A laptop might be about 1 Gflop; again, anything reasonable.

(g) An *Economist* article on semiconductor manufacturing (12/18/19) says "These chips will cram more than 170 million transistors onto each square millimetre of silicon, creating structures with [feature sizes] as small as 5 nanometres."

(i) If one of these new chips is one square centimeter, how many transistors could it hold.

17 billion. There are 100 sq mm in one sq cm (not 10!) so it's a factor of 100.

(ii) When I built a chip in 1980, feature sizes were 5 micrometers. All else being equal, how many transistors might have fit onto a square millimeter of my chip?

170. Feature sizes were 1000 times bigger so the area of a transistor is 1,000,000 times bigger. Almost no one got this part.

(h) An article about the NYC marathon says "Every competitor will wear a shoe with a chip that will record their progress, and can send e-mail updates every five kilometers to spectators who subscribe to the service." For each of the following inferences that a non-technical reader might make from this quotation, assess whether they are likely to be correct or unlikely to be correct.

the chip has enough memory to store at least a dozen time measurements	**likely**	**unlikely**
the chip uniquely identifies the runner who wears it	**likely**	**unlikely**
the chip uses GPS to determine how far the runner has run so far	**likely**	**unlikely**
the chip sends e-mail messages to a server	**likely**	**unlikely**
the maximum number of e-mail subscribers depends on properties of the chip	**likely**	**unlikely**

The chip is basically just a prox, so it can't do anything except report its number when it goes near a sensor.

(i) The hex value **`FF00FF`** can be interpreted as an RGB color. Suppose that instead this value is interpreted simply as a 24-bit integer, stored in a variable **`v`**, and incremented by 1 in a Javascript statement like **`v = v+1`**.

(i) What is the resulting value in hexadecimal?

`FF0100`.

(ii) What color is the resulting value closest to?

red **green** **blue** **yellow** **cyan** **magenta** **black** **white**

red

(j) Princeton has two blocks of IPv4 addresses, 140.180.x.x and 128.112.x.x.

(i) How many Princeton IPv4 addresses are there in total? An expression is fine.

2^17. Two subnets, each of size 2^16. We did this in class one day if memory serves.

(ii) What is the numerically largest possible Princeton IPv4 address, expressed in dotted decimal?

140.180.255.255. As noted many times, each number is a decimal representation of an 8-bit number, so 255 is the maximum value.

(k) "But you can't look up all those license numbers in time," Drake objected. "We don't have to, Paul. We merely …" (*Perry Mason: The Case of the Angry Mourner*, a 1951 pot-boiler by Erle Stanley Gardner). Suppose you have ten million license numbers and you have to find all the duplicates in a hurry.

(i) *In a dozen words at most*, describe an efficient algorithm for finding all duplicate license numbers in a list of ten million.

Sort, look for adjacent duplicates.

(ii) How does the running time of your algorithm depend on **N**, the number of items in the list?

n log n

(l) ASCII is technically an 8-bit code but most of the time only the rightmost 7 bits are used and the leftmost bit is always zero. Suppose that instead we use this leftmost bit as a parity bit.

(i) The hex value of the ASCII character **`z`** without parity is **`7A`**. What is its hex value with ***odd*** parity?

7A. Parity doesn't change this one.

(ii) The hex value of the ASCII character **?** without parity is **3F**. What is its hex value with ***odd*** parity?

BF. Parity sets the leftmost bit to 1.

(iii) The hex value **AA** is an ASCII character with ***even*** parity. What character is it?

*****. Removing the parity bit leaves 2A. This part seemed to be hard.

(m) If I use my phone to send mail to a friend in England, as the mail goes from me to him, which of these mechanisms (A) is almost sure to be used? (B) might be used but need not be? (C) is very unlikely to be used? Circle the best answer.

TCP/IP	**almost sure**	**might be**	**very unlikely**
traceroute	**almost sure**	**might be**	**very unlikely**
fiber-optic cable	**almost sure**	**might be**	**very unlikely**
IPv6	**almost sure**	**might be**	**very unlikely**
wired Ethernet	**almost sure**	**might be**	**very unlikely**

(n) A *NY Times* article about E-ZPass, the electronic highway toll system, says, "A list of valid and invalid tag numbers is sent every day to computer drives in every toll booth. As a vehicle drives through an E-ZPass lane, a high-speed optical reader almost instantly identifies the tag mounted to a dashboard or windshield and matches it against the list to see if the holder has enough money set aside to pay the toll." Identify three technical "facts" in this quotation that are almost surely wrong, or at least badly misleading. ***Be brief*** – a few well-chosen words should be adequate for each.

There are no drives in toll booths, there is no daily list, there is no immediate check against the list, there's no checking money, the devices are radio, not optical. Terrible reporting. The question wasn't as clear as it should have been, so graded generously.

(o) A *CNET* blog post (1/8/20) says that there were 100 million Alexas at the start of 2019 and 200 million at the start of 2020. Assuming (most improbably!) that this represents a smooth exponential growth that will continue into the future,

(i) What is the rate of increase of the number of Alexas each month?

6%. The number doubles in 12 months.

(ii) During what year will the number of Alexas first exceed the number of people in the world?

2025. It gets to 6.4B at the beginning of 2025 and will be at 12.8B at the end. I accepted 2026 with a small penalty.

(p) IPv6 packets contain an 8-bit field called the "hop limit". Each router that receives a packet decreases the hop limit field by 1. Normally the packet is then forwarded, but if the hop limit becomes zero, the packet is discarded and an error packet is sent back to the source.

(i) What is the maximum number of routers that a packet could possibly reach before it is discarded?

255. I accepted 256 since it's bit ambiguous about when things happen.

(ii) Traceroute works by a cute hack. It sends out a packet with a hop limit of 1, and records the IP address of the router that discarded the packet. It then sends out a package with a hop limit of 2 and records that router, which is the second in the path. It repeats this process until the packet ultimately reaches its destination. How does the total number of hops taken by all these packets grow as a function of **n**, the length of the path to the ultimate destination?

logarithmic **linear** **n log n** **quadratic** **exponential** **can't tell**

quadratic. The number of hops is 1 + 2 + 3 + … This part wasn't well handled; few saw it as quadratic.

(q) Refer to the ASCII chart on the cover page of the exam.

(i) By what *decimal* numeric value does a lower case ASCII letter differ from its corresponding upper case value?

32. Each table row has 16 values and corresponding letters are separated by two rows!

(ii) If **`ch`** is a variable that contains an arbitrary ASCII character, explain in at most half a dozen words what this test is trying to determine. Do NOT just repeat the code in words.

```
if ch >= 48 and ch <= 57 ...
```

Is ch a digit? Remember that row and column labels are hex, not decimal.

(r) In managing their nefarious activities, Eve and Mallory naturally use public-key cryptography to exchange encrypted email. Suppose that Alice learns Eve's private key. What can Alice now do?

Alice can convince Mallory that she (Alice) is really Eve	**true**	**false**
Alice can convince Trent that she is really Eve	**true**	**false**
Alice can convince Eve that she (Alice) is really Mallory	**true**	**false**
Alice can read an encrypted message from Mallory to Eve	**true**	**false**
Alice can read an encrypted message from Eve to Mallory	**true**	**false**

If Alice knows Eve's private key, she is Eve for all cryptographic purposes.

(s) [10 pts] Random quickies: Circle the best answers.

Machine learning algorithms for spam detection are often based on supervised learning	**true**	**false**
Trans-oceanic Internet traffic is transmitted through communications satellites	**true**	**false**
Root servers exchange their traffic at Internet Exchange Points	**true**	**false**
My browser will use public key crypto when I buy a book at Amazon	**true**	**false**
My browser will use secret key crypto when I buy a book at Amazon	**true**	**false**
"An IP address is like a zip code: it tells where your computer is located"	**true**	**false**
New top-level domains like **`.info`** are created by registrars like GoDaddy	**true**	**false**
The comment `/* You are not expected to understand this */` comes from Unix source code	**true**	**false**
The Luhn algorithm guarantees that a credit card number is valid	**true**	**false**
Google competes with DoubleClick for advertising revenue	**true**	**false**